SYSTEMS, METHODS, AND MEDIA FOR ON-LINE BOOSTING OF A CLASSIFIER

Applicant: Arizona Board of Regents, a body corporate of the State of Arizona, acting for and on behalf of, Arizona State University, Scottsdale, AZ (US)

Inventors: Nima Tajbakhsh, Tempe, AZ (US); Hong Wu, Tempe, AZ (US); Wenzhe Xue, Phoenix, AZ (US); Jianming Liang, Scottsdale, AZ (US)

Assignee: Arizona Board of Regents, a body corporate of the State of Arizona, acting for and on behalf of, Arizona State University, Scottsdale, AZ (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 86 days.

Filed: Sep. 17, 2012

Prior Publication Data

Related U.S. Application Data
Provisional application No. 61/535,933, filed on Sep. 16, 2011.

Int. Cl.
G06K 9/00 (2006.01)
G06K 9/62 (2006.01)

U.S. Cl.
CPC G06K 9/62 (2013.01)

Field of Classification Search
None

See application file for complete search history.

Abstract
Systems, methods, and media for on-line boosting of a classifier are provided, comprising: receiving a training sample; for each of a plurality of features, determining a feature value for the training sample and the feature, using the feature value to update a histogram, and determining a threshold for a classifier of the feature; for each of the plurality of features, classifying the training sample using the threshold for the classifier of the feature and calculating an error associated with the classifier; selecting a plurality of best classifiers from the classifiers; and, for each of the plurality of best classifiers, assigning a voting weight to the one of the plurality of best classifiers.

21 Claims, 4 Drawing Sheets
References Cited

U.S. PATENT DOCUMENTS

2008/0260230 A1 10/2008 Gotardo et al. ................. 382/130
2009/0175515 A1 7/2009 Schummers ......... 382/130
2010/0061615 A1 2/2010 Von Borg et al. .... 382/130
2010/0098308 A1 4/2010 Lakare et al. .............. 382/130
2010/0177944 A1 7/2010 Madhabushi et al. ... 382/130
2011/0191283 A1 8/2011 Voigt et al. .............. 382/159

OTHER PUBLICATIONS


(56) References Cited

OTHER PUBLICATIONS


Liang, J. and Bi, J., “Computer Aided Detection of Pulmonary Embolism with Tobogifying and Multiple Instance Classification in CT Pulmonary Angiography”, in Proceedings of the 20th Int'l Conference of Information Processing in Medical Imaging (Kerkadri, NL, Jul. 2-6, 2007, pp. 630-641.


References Cited


* cited by examiner
Input:
Training samples $S = (x_t, y_t), y_t \in \{1, -1\}, t = 1, 2, ..., T$
A feature pool containing $M$ features
Number of features to be selected, $N$ ($N < M$)
Maximum number of classification attempts, $N_{\text{att}}$

Initialize:
initialize each weak learner: $\lambda_m^{\text{corr}} = 1$ and $\lambda_m^{\text{wrong}} = 1$

Learning process
for $t = 1, 2, ..., T$

[Step 0: Initialize parameters]
$\lambda = 1$ {Initialize the importance weight}
i = 1 {Initialize the number of attempts}
$\Phi = []$ {Initialize the set of selected weak learners}

while $i \leq N_{\text{att}}$

[Step 1: Update all weak learners]
for $m = 1, 2, ..., M$
$WL_m = \text{update}(x_t, y_t, WL_m)$
end for

[Step 2: Select the best $N$ features]
for $n = 1, 2, ..., N$

[Step 2.1: Reward and Punish each weak learner]
for $m = 1, 2, ..., M$
if $WL_m(x_t) = y_t$
$\lambda_m^{\text{corr}} = \lambda_m^{\text{corr}} + \lambda$ {correct decision}
else
$\lambda_m^{\text{wrong}} = \lambda_m^{\text{wrong}} + \lambda$ {wrong decision}
end if
end for

[Step 2.2: Select the best weak learner]
$m^* = \arg \min_{m \in \Phi} \text{error}_m$
Add $m^*$ to set $\Phi$

[Step 3: Calculate the voting weight]
$\alpha_n = \frac{1}{2}(\text{error}_m)$

[Step 4: Update $\lambda$]
if $WL_m(x_t) = y_t$
$\lambda = \frac{\lambda}{2(1 - \text{error}_m)}$ {correct decision}
else
$\lambda = \frac{\lambda}{2(\text{error}_m)}$ {wrong decision}
end if
end for

[Step 3: Classify after updating]
$F(x_t) = \text{sign}(\sum_{j \in \Phi} \alpha_j \times WL_j(x_t))$
if $F(x_t) = y_t$
break {no further attempts}
else
$i = i + 1$
end if
end while
end for

Output
Final classifier $F(x) = \text{sign}(\sum_{j \in \Phi} \alpha_j \times WL_j(x))$
FIG. 4

400

Memory

404

Hardware
Processor
(e.g.,
microprocessor,
digital signal
processor, etc.)

402

Communications
Bus

414

Database and/or
Storage

412

User Input Device

410

Display Interface
and Display

408

External
Devices

406

Communication
Interface(s)

408
SYSTEMS, METHODS, AND MEDIA FOR ON-LINE BOOSTING OF A CLASSIFIER

CROSS REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application No. 61/535,933, filed Sep. 16, 2012, which is hereby incorporated by reference herein in its entirety.

TECHNICAL FIELD

The disclosed subject matter relates to systems, methods, and media for on-line boosting of a classifier.

BACKGROUND

Automatic detection of certain content in images and/or other forms of data is of ever-increasing importance for machine vision, security, computer-aided diagnosis and other applications. For example, automated detection of anatomic structures is an important functionality for navigating through large 3D image datasets and supporting computer-aided diagnosis (CAD).

A classifier is a mechanism that can be used to perform automatic detection in such applications. Once trained, a classifier can indicate whether an image includes a certain object, such as an anatomic structure. Based on the amount of training, a classifier can exhibit better or worse performance. With an off-line classifier, training must be done in advance of normal use of the classifier, while with an on-line classifier, training can be done concurrently with normal use of the classifier (which training is known as on-line boosting of the classifier). Because of this ability to train, during normal use, and hence continuously improve performance while being used, on-line classifiers are increasing in popularity.

However, known on-line classifiers suffer from many deficiencies. Accordingly, new on-line classifiers are desirable.

SUMMARY

Systems, methods, and media for on-line boosting of a classifier are provided.

In accordance with some embodiments, systems for on-line boosting of a classifier are provided. The systems comprising: a hardware processor that is configured to receive a training sample; for each of a plurality of features, determine a feature value for the training sample and the feature, use the feature value to update a histogram, and determine a threshold for a classifier of the feature; and, for each of the plurality of best classifiers, assign a voting weight to the one of the plurality of best classifiers.

In accordance with some embodiments, a non-transitory computer-readable media containing computer-executable instructions that, when executed by a processor, cause the processor to perform a method for on-line boosting of a classifier are provided. The method comprises: receiving a training sample; for each of a plurality of features, determining a feature value for the training sample and the feature, using the feature value to update a histogram, and determining a threshold for a classifier of the feature; and, for each of the plurality of features, classifying the training sample using the threshold for the classifier of the feature and calculating an error associated with the classifier; selecting, using the hardware processor, a plurality of best classifiers from the classifiers; and, for each of the plurality of best classifiers, assigning a voting weight to the one of the plurality of best classifiers using the hardware processor.

In accordance with some embodiments, a non-transitory computer-readable media containing computer-executable instructions that, when executed by a processor, cause the processor to perform a method for on-line boosting of a classifier are provided. The method comprises: receiving a training sample; for each of a plurality of features, determining a feature value for the training sample and the feature, using the feature value to update a histogram, and determining a threshold for a classifier of the feature; and, for each of the plurality of features, classifying the training sample using the threshold for the classifier of the feature and calculating an error associated with the classifier; selecting, using the hardware processor, a plurality of best classifiers from the classifiers; and, for each of the plurality of best classifiers, assigning a voting weight to the one of the plurality of best classifiers.

BRIEF DESCRIPTION OF THE DRAWINGS

The above and other objects and advantages of the invention will be apparent upon consideration of the following detailed description, taken in conjunction with the accompanying drawings, in which like reference characters refer to like parts throughout, and in which:

FIG. 1 is an example of a block diagram of a system for classifying images in accordance with some embodiments of the disclosed subject matter;

FIG. 2 is an example of a flow diagram of a process for updating a classifier in accordance with some embodiments of the disclosed subject matter;

FIG. 3 is an example of a flow diagram of a process for updating a classifier in accordance with some embodiments of the disclosed subject matter; and

FIG. 4 is an example of a block diagram of hardware for a system for classifying images in accordance with some embodiments of the disclosed subject matter.

DETAILED DESCRIPTION

In describing exemplary embodiments of the disclosed, subject matter illustrated in the drawings, specific terminology may be employed for the sake of clarity. However, the disclosed subject matter is not intended to be limited to the specific terminology so selected, and it is to be understood that each specific element includes all technical equivalents which operate in a similar manner.

Systems, methods, and media for on-line boosting of a classifier are provided.

FIG. 1 illustrates an example of a system 100 that performs classification in accordance with some embodiments. In some embodiments, system 100 can be a computer aided detection system. As illustrated, an input image 102 (such as a medical imaging image of an anatomic structure) can be received and provided to a detector 104. Detector 104 can be a system including a hardware processor that implements a linear classifier of the form:

\[ F(x) = \text{sgn} \left( \sum_{j=1}^{J} \alpha_j \Phi(x) W_{L_j}(x) \right) \]

where \( W_{L_j}(x) \) is an output returned by the \( j \)th weak learner for image \( x \), \( \alpha_j \) is a voting weight of the \( j \)th weak learner, and \( \Phi \) denotes the set containing the indices of selected weak learners. This linear classifier can be viewed as a classifier that implements a linear classifier of the form:

\[ \sum_{j=1}^{J} \alpha_j \Phi(x) W_{L_j}(x) \]

where \( W_{L_j}(x) \) is an output returned by the \( j \)th weak learner for image \( x \), \( \alpha_j \) is a voting weight of the \( j \)th weak learner, and \( \Phi \) denotes the set containing the indices of selected weak learners. This linear classifier can be configured to detect a certain anatomic structure in input image 102 by...
having been “trained” as described herein. The classifier can then output a “1” if the sample contains the desired certain anatomie structure, otherwise it can return a “−1”.

Because anatomie structures may appear at different scales, the detector may perform a multi-scale search to locate the anatomic structures.

Image 102 and its classification (i.e., “1” or “−1”) can then be reviewed by a verifier 106. Verifier 106 can be any suitable mechanism for authoritatively determining whether image 102 was classified correctly by detector 104 and for specifying a “ground truth” for a desired object (i.e., a true location for the desired object). For example, verifier 106 can be a device through which a radiologist inspects the image and determines whether the certain anatomie structure is in fact present and therefore whether the classification is correct. Images that are correctly classified by detector 104 can be output at images 108. Images that are incorrectly classified by detector 104 can be output as images 110. Images 110 can include the authoritative classification information (i.e., information that indicates the true classification of the images) and the images and information can be provided to an updatie 112 for updating the linear classifier in detector 104. Updater 112 can update the linear classifier in any suitable manner, such as that described herein in connection with FIGS. 2 and 3.

In accordance with some embodiments, these mechanisms can be used for automatic detection of anatomic structures using on-line boosting. For example, in some embodiments, these mechanisms can be used to detect a human pulmonary trunk, a human carina, and a human aortic arch.

Although system 100 and the mechanisms of FIGS. 2, 3, and 4 are described herein in the context of detection of anatomic structures in medical imaging images, it should be apparent to one of ordinary skill in the art that this is just an example of an application of this technology and that this technology can be used for any suitable application. For example, in some embodiments, this technology can be used to classify and/or perform detection in any suitable image or in any suitable data, can be used for real-time object detection and/or real-time object tracking, can be used for intelligent video surveillance, content based image retrieval, face and activity recognition, traffic control, and human-computer interfaces, etc.

As used herein, the term “image” may refer to multi-dimensional data composed of discrete image elements (e.g., pixels for two-dimensional images and voxels for three-dimensional images). The images may be, for example, a medical image of a subject collected using a computer tomography system, a magnetic resonance imaging system, an ultrasound imaging system, or any other medical imaging system or imaging modality known to one of skill in the art. The images may also be provided from non-medical contexts, such as, for example, remote sensing systems, electron microscopy systems, etc. The methods of the disclosed subject matter are not limited to such images, and can be applied to images of any dimension, e.g., a two-dimensional picture, a three-dimensional volume, or a four-dimensional space. For a two-dimensional or three-dimensional image, the domain of the image is typically a two-dimensional or three-dimensional rectangular array, where each pixel or voxel can be addressed with reference to a set of two or three mutually orthogonal axes.

Turning to FIG. 2, a flow diagram of a process 200 for training a classifier that can be used in some embodiments is shown. As illustrated, after process 200 begins at 202, the process can select a pool of features to be used for classification and set-up corresponding weak learners for the features. Any suitable features can be used in some embodiments. For example, in some embodiments, 2D Haar patterns can be used as features for classification. More particularly, for example, in some embodiments, four 2D Haar patterns at different positions, scales, and aspect ratios (and/or any other variations) can be used to form an initial set of 101,400 (or any other suitable number) features, and from this set, 1000 (or any other suitable number) features can be randomly selected to form a pool of features. As another example, in some embodiments, 3D Haar patterns, local binary patterns (LBP), histogram of gradients (HOG), SIFT, or any other suitable patterns, can be used.

Each feature in the pool can then be assigned a weak learner. Any suitable weak learner can be used in some embodiments. For example, in some embodiments, a weak learner can be implemented using a decision stump that compares a feature value for a sample and the feature to a threshold for the feature that is dynamically set based on two histograms for the feature: one for positive samples; and one for negative samples.

Any suitable feature can be used and the feature value can be calculated in any suitable manner. For example, in some embodiments, when using 2D Haar patterns, a feature value for a feature can be calculated as the sum of the intensities of all sample pixels corresponding to white pixels in the Haar pattern for the feature minus the sum of all sample pixels corresponding to black pixels in the Haar pattern for the feature.

Because samples can be processed sequentially as described below, the two histograms can be incrementally assembled over time in some embodiments. To form the set of bins in the histograms before processing all samples, a range of the feature values for the samples for each feature can be estimated in advance by examining feature values computed from a subset of the samples in some embodiments. Any suitable number of samples can be included in this subset in some embodiments. This subset set can subsequently be discarded and not used in the training the classifier in some embodiments. The estimated range for each feature can then be divided (e.g., equally) into 100 (or any other suitable number) of bins.

Next, the process can receive a training sample at 206. This training, sample can be received in any suitable manner and from any suitable source for any suitable purpose in any suitable format. For example, in some embodiments, such a training sample can be received from a file, from memory, from a computer network, from a scanner, from a video sensor, from a display or display driver, etc. As another example, in some embodiments, when used for detection of anatomic structures, the training sample can be from medical imaging data, such as computed tomography (CT) scan images or any other suitable medical imaging scan images. As still another example, such sample can be received in at any suitable one or more resolutions, in color, in grayscale, etc. From such data, positive and negative samples can be extracted and resized to any suitably sized sub-images (such as 25×25 pixel images, for example). Any suitable number of positive and negative samples (such as 4000 each) can be used in some embodiments. The positive samples can be selected at, and/or up to some distance away (e.g., such as 5 pixels away) from the ground truth (i.e., a known-accurate center point) of desired structures, and negative samples can be randomly selected such that the desired structures are not included, in some embodiments.

At 208, histograms of the weak learners can then be updated based on the training sample and new decision thresholds can be calculated, in some embodiments. More particularly, for example, for each feature in the pool of
features, when a training sample arrives, the feature’s value for the sample can be calculated, and, depending on the sample’s label as being a positive sample or a negative sample, the pertinent histogram can be selected, the bin to which the sample belongs can be determined, and the frequency of the associated bin can be updated, in some embodiments. Any training samples whose feature values fall beyond the lower range or the upper range of the bin can be assigned to the first bin or the last bin, respectively, in some embodiments. Once the histograms are updated, each feature’s decision threshold can be defined as the intersection of the positive and negative histograms for that feature. These thresholds are chosen among the histograms bins such that maximum classification rate is obtained.

Each updated weak learner can then classify the sample and be rewarded or punished with the sample’s importance weight \( \lambda \) for a correct or wrong classification, respectively, at 210, in some embodiments. The importance weight \( \lambda \) can initially be set to 1 and can subsequently be updated during the learning process at 216 as described below. The rewards and punishments a weak learner receives can be accumulated in \( \lambda_{\text{true}} \) and \( \lambda_{\text{false}} \), respectively, and \( \lambda_{\text{true}} \) and \( \lambda_{\text{false}} \) can be used to calculate the error rate of each weak learner as follows:

\[
\text{error}_n = \frac{\lambda_{\text{false}}}{\lambda_{\text{true}} + \lambda_{\text{false}}},
\]

where \( n \) is an index for the weak learner.

Next, at 212, process 200 can select the best weak learner that has not been previously selected as a best weak learner for the present sample, in some embodiments. This best weak learner can be selected as the weak learner having the least error rate. In some embodiments, this best weak learner can be selected as the weak learner having the best Gini index, entropy, sensitivity, specificity, and/or any other characteristic. The selected weak learner can then be added to the set \( \Phi \) that contains the best weak learners selected so far. Any suitable number of best weak learners can be selected in some embodiments. For example, in some embodiments, 100 best weak learners can be selected.

Once a best weak learner has been chosen, its corresponding voting weight can be computed at 214, in some embodiments. The voting weight can impact the extent to which the selected best weak learner contributes to the final classifier (which is a linear classifier that is a weighted combination of the best weak learners). The voting weight \( \alpha_n \) (\( n \mapsto N \)), corresponding to the \( n^{th} \) best weak learner, can be given by

\[
\alpha_n = \frac{1}{2} \ln \left( \frac{1 - \text{error}_n}{\text{error}_n} \right).
\]

where \( m^* \) denotes the best weak learner index (\( m^* \mapsto M \)).

Next, according to the classification outcome and error rate of the selected best weak learner, the importance weight \( \lambda \) can be updated. The importance weight can increase for a wrong classification and decrease for a correct classification. For example, for a wrong classification, the importance weight \( \lambda \) can be updated as:

\[
\lambda = \frac{\lambda}{2(\text{error}_n)}
\]

and for a correct classification, the importance weight \( \lambda \) can be updated as:

\[
\lambda = \frac{\lambda}{2(1 - \text{error}_n)}
\]

in some embodiments.

After the importance \( \lambda_{\text{weight}} \) has been updated at 216, process 200 can determine at 218 whether another best weak learner is to be selected. This determination can be made on any suitable basis. For example, this determination can be made based on a count (such as 100) of best weak learners selected for the present sample so far. If another best weak learner is to be selected, process 200 can branch back to 212.

Otherwise, process 200 can update the final linear classifier at 220 using the selected \( N \) best weak learners in some embodiments. For example, in some embodiments, the final linear classifier can be defined as:

\[
F(x) = \text{sign} \left[ \sum_{\alpha \in \Phi} \alpha \cdot WL_n(x) \right]
\]

where \( WL_n(x) \) is the output returned by the \( n^{th} \) best weak learner for sample \( x \), \( \alpha \) is the voting weight of the \( n^{th} \) best weak learner, and \( \Phi \) denotes the set containing the indices of selected best weak learners. The classifier outputs a ‘1’ if the sample contains the desired object, otherwise it returns a ‘-1’.

Next, at 222, process 200 can classify the present sample using the final linear classifier in some embodiments. At 224, the process can determine whether the present sample was classified correctly in some embodiments. If so, then process 200 can loop back to 206 to receive the next training sample. Otherwise, process 200 can branch to 226 to determine if a final classification attempt limit has been met for the present sample. Any suitable attempt limit, such as five (or any other suitable number) attempts, can be used in some embodiments. If the attempt limit has not been met, then process 200 can loop back to 208. Otherwise, the process can loop back to 206.

It should be understood that the above steps of the flow diagrams of FIG. 2 may be executed or performed in any order or sequence not limited to the order and sequence shown and described in the figure. Also, some of the above steps of the flow diagrams of FIG. 2 may be executed or performed substantially simultaneously where appropriate or in parallel to reduce latency and processing times.

FIG. 3 shows an example of pseudo-code corresponding to portions of process 200 of FIG. 2 in accordance with some embodiments.

In accordance with some embodiments, any suitable hardware and/or software can be used to perform the mechanisms described herein (such as those illustrated in, and described in connection with, FIGS. 1, 2, and 3). For example, a general purpose device such as a computer or a special purpose device such as a client, a server, etc. can be used to execute software for performing the mechanisms described herein. Any of these general or special purpose devices, such as device 400 of FIG. 4, can include any suitable components such as a hardware processor 402 (which can be a microprocessor, digital signal processor, a controller, etc.), memory 404, communication interfaces 406, a display interface and display 408, user input devices 410, a database and/or storage 412, a communications bus 414, etc. Communications interfaces
can enable the hardware and/or software to communicate with other communications networks (e.g., such as the Internet, wired networks, wireless networks, etc.), other equipment (e.g., such as medical scanning (e.g., such as a computed tomography scanner) diagnosis, display, etc. equipment), and/or any other suitable networks, devices, etc. This hardware and/or software can be implemented as part of other equipment (e.g., such as medical scanning (e.g., such as a computer tomography scanner), diagnosis, display, etc. equipment) or can be implemented as stand-alone equipment (which can be coupled to other equipment).

In some embodiments, any suitable computer readable media can be used for storing instructions for performing the processes described herein. For example, in some embodiments, computer readable media can be transitory or non-transitory. For example, non-transitory computer readable media can include media such as magnetic media (such as hard disks, floppy disks, etc.), optical media (such as compact discs, digital video discs, Bi-ray discs, etc.), semiconductor media (such as flash memory, electrically programmable read only memory (EPROM), electrically erasable programmable read only memory (EEPROM), etc.), any suitable media that is not fleeting or devoid of any semblance of permanence during transmission, and/or any suitable tangible media. As another example, transitory computer readable media can include signals on networks, in wires, conductors, optical fibers, circuits, any suitable media that is fleeting and devoid of any semblance of permanence during transmission, and/or any suitable intangible media.

Although the invention has been described and illustrated in the foregoing illustrative embodiments, it is understood that the present disclosure has been made only by way of example, and that numerous changes in the details of implementation of the invention can be made without departing from the spirit and scope of the invention, which is limited only by the claims which follow. Features of the disclosed embodiments can be combined and rearranged in various ways.

What is claimed is:

1. A system for on-line boosting of a classifier, comprising:
   a hardware processor that is configured to:
   access a selection of a plurality of image features for the classifier;
   receive an online training sample not used to select the plurality of image features;
   for each of the plurality of image features:
   determine a feature value for the online training sample and the feature using the online training sample;
   determine one of a plurality of bins of a histogram that corresponds to the online training sample based on the feature value for the online training sample, wherein:
   when the feature value falls below a lower range of the plurality of bins, a lowest bin in the plurality of bins is determined to be the one of the plurality of bins; and
   when the feature value falls above an upper range of the set of bins, a highest bin in the plurality of bins is determined to be the one of the plurality of bins;
   use the feature value to update the one of the plurality of bins of the histogram; and
   determine a threshold for a classifier of the feature;
   for each of the plurality of image features, classify the training sample using the threshold for the classifier of the feature and calculate an error associated with the classifier;
   select a plurality of best classifiers from the classifiers, wherein each of the plurality of best classifiers have not been previously selected as a best classifier; and
   for each of the plurality of best classifiers, assign a voting weight to the one of the plurality of best classifiers that contributes to a final classifier, wherein the final classifier is a linear classifier that is a weighted combination of the plurality of best classifiers.

2. The system of claim 1, wherein the hardware processor is also configured to use the plurality of best classifiers and the voting weights to perform classification.

3. The system of claim 1, wherein the classification is performed using a linear classifier.

4. The system of claim 1, wherein the hardware processor is also configured to:
   calculate an importance weight for the classifier of one of the plurality of features; and
   use the importance weight to calculate the error of the classifier of another of the plurality of features.

5. The system of claim 1, wherein the training sample is a medical imaging image.

6. The system of claim 5, wherein the medical imaging image is a computed tomography image.

7. The system of claim 1, wherein each of a plurality of features is a Haar pattern.

8. A method for on-line boosting of a classifier, comprising:
   accessing a selection of a plurality of image features for the classifier;
   receiving an online training sample using a hardware processor not used to select the plurality of image features;
   for each of the plurality of image features, using the hardware processor:
   determining a feature value for the online training sample and the feature using the online training sample;
   determining one of a plurality of bins of a histogram that corresponds to the online training sample based on the feature value for the online training sample, wherein:
   when the feature value falls below a lower range of the plurality of bins, a lowest bin in the plurality of bins is determined to be the one of the plurality of bins; and
   when the feature value falls above an upper range of the set of bins, a highest bin in the plurality of bins is determined to be the one of the plurality of bins;
   using the feature value to update the one of the plurality of bins of the histogram; and
   determining a threshold for a classifier of the feature for each of the plurality of image features, classify the training sample using the threshold for the classifier of the feature and calculate an error associated with the classifier;
   select a plurality of best classifiers from the classifiers, wherein each of the plurality of best classifiers have not been previously selected as a best classifier; and
   for each of the plurality of best classifiers, assign a voting weight to the one of the plurality of best classifiers that contributes to a final classifier, wherein the final classifier is a linear classifier that is a weighted combination of the plurality of best classifiers.
9. The method of claim 8, further comprising using the plurality of best classifiers and the voting weights to perform classification using the hardware processor.

10. The method of claim 8, wherein the classification is performed using a linear classifier.

11. The method of claim 8, further comprising:
calculating an importance weight for the classifier of one of the plurality of features using the hardware processor; and
using the importance weight to calculate the error of the classifier of another of the plurality of features using the hardware processor.

12. The method of claim 8, wherein the training sample is a medical imaging image.

13. The method of claim 12, wherein the medical imaging image is a computed tomography image.

14. The method of claim 8, wherein each of a plurality of features is a Haar pattern.

15. A non-transitory computer-readable medium containing computer-executable instructions that, when executed by a processor, cause the processor to perform a method for on-line boosting of a classifier, the method comprising:
accessing a selection of a plurality of image features for the classifier;
retrieving an online training sample not used to select the plurality of image features;
for each of the plurality of image features:
determining a feature value for the online training sample and the feature using the online training sample;
determining one of a plurality of bins of a histogram that corresponds to the online training sample based on the feature value for the online training sample, wherein:
when the feature value falls below a lower range of the plurality of bins, a lowest bin in the plurality of bins is determined to be the one of the plurality of bins; and
when the feature value falls above an upper range of the set of bins, a highest bin in the plurality of bins is determined to be the one of the plurality of bins;
using the feature value to update the one of the plurality of bins of the histogram; and
determining a threshold for a classifier of the feature for each of the plurality of image features, classifying the training sample using the threshold for the classifier of the feature and calculating an error associated with the classifier;
selecting a plurality of best classifiers from the classifiers, wherein each of the plurality of best classifiers have not been previously selected as a best classifier; and
for each of the plurality of best classifiers, assigning a voting weight to the one of the plurality of best classifiers that contributes to a final classifier, wherein the final classifier is a linear classifier that is a weighted combination of the plurality of best classifiers.

16. The non-transitory computer-readable medium of claim 15, wherein the method further comprises using the plurality of best classifiers and the voting weights to perform classification.

17. The non-transitory computer-readable medium of claim 15, wherein the classification is performed using a linear classifier.

18. The non-transitory computer-readable medium of claim 15, wherein the method further comprises:
calculating an importance weight for the classifier of one of the plurality of features; and
using the importance weight to calculate the error of the classifier of another of the plurality of features.

19. The non-transitory computer-readable medium of claim 15, wherein the training sample is a medical imaging image.

20. The non-transitory computer-readable medium of claim 19, wherein the medical imaging image is a computed tomography image.

21. The non-transitory computer-readable medium of claim 15, wherein each of a plurality of features is a Haar pattern.